《数理统计分析》

实验报告

|  |  |
| --- | --- |
| 实验名称 | 数据处理 |
| 班 级 | 区块链221 |
| 学生姓名 | 罗彧 |
| 学生学号 | 2022131013 |
| 指导教师 | 高路路 |
| 成 绩 |  |

2024 年 11月19 日 成都信息工程大学 人工智能学院

**一、实验目的：**

学习、理解基于R语言数据处理和分析的基本操作，和对该内容的具体实践。

**二、实验环境：**

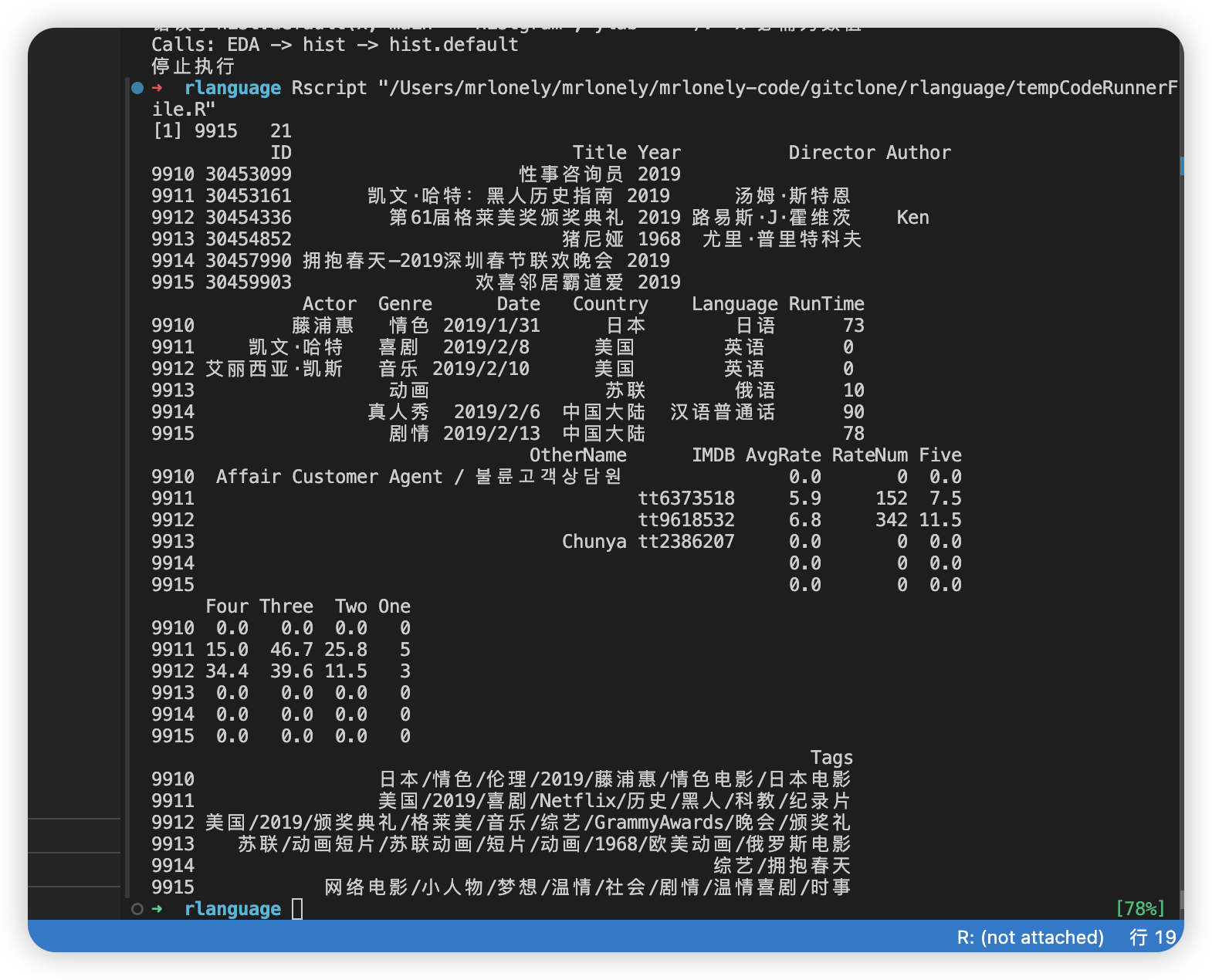
Windows10操作系统、R、RStudio

**三、实验步骤：**

1. 根据以下要求，针对上次实验生成的电影数据集（电影编号和学号结尾两位数字相同），对数据进行处理和分析：

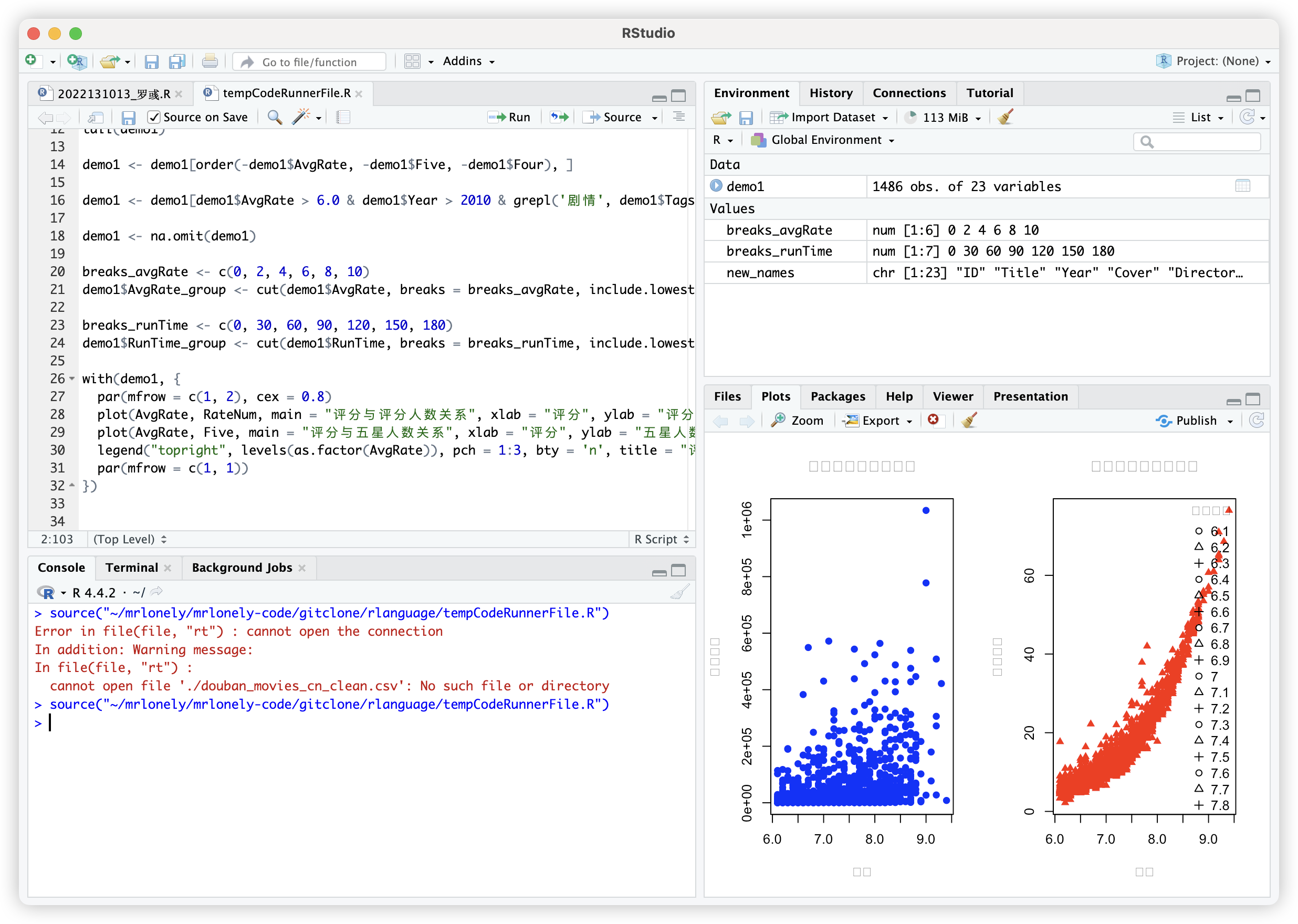
文本

描述已自动生成1）. 使用下面的函数，观察评分和时长变量数据性质

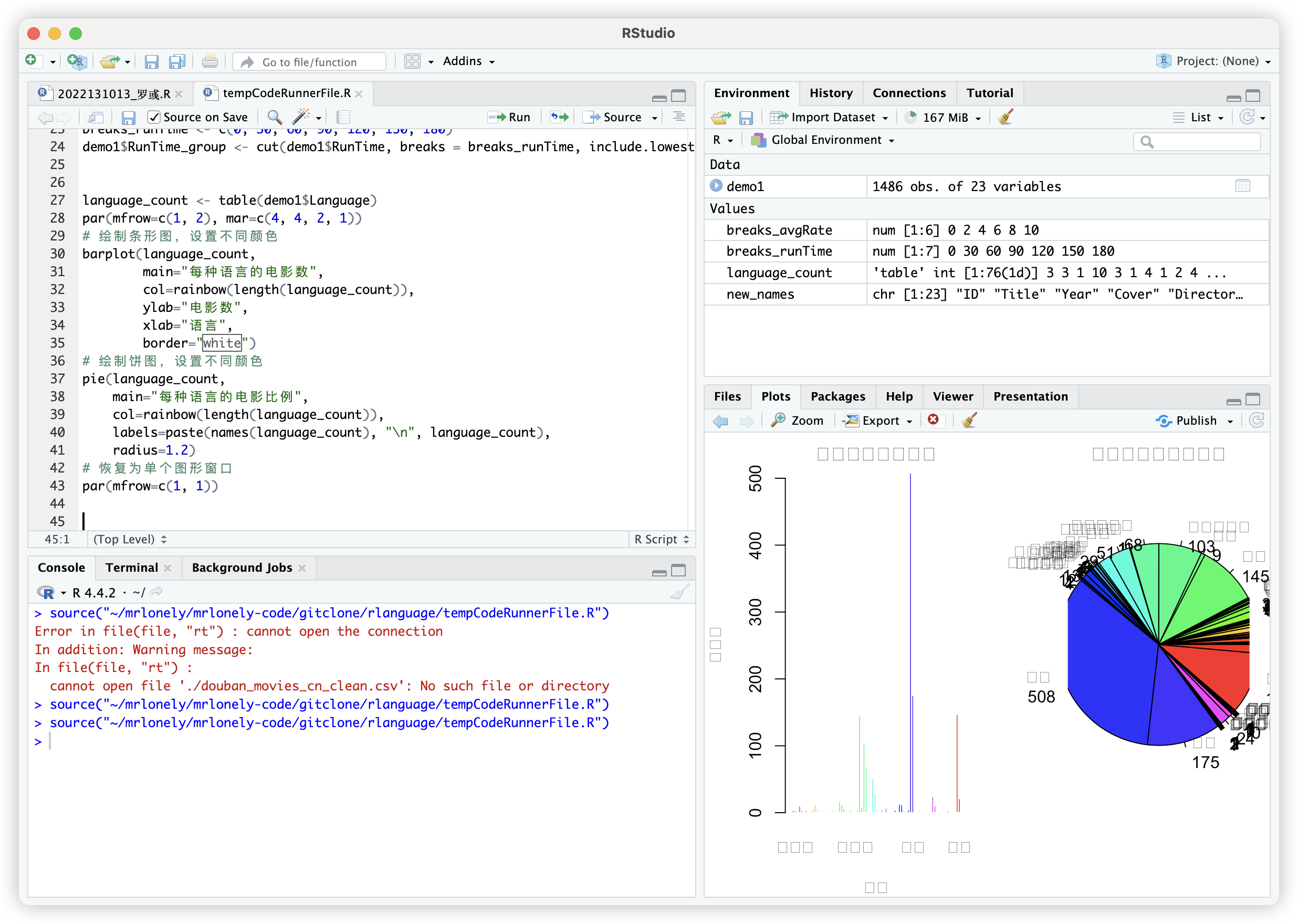


图表, 散点图

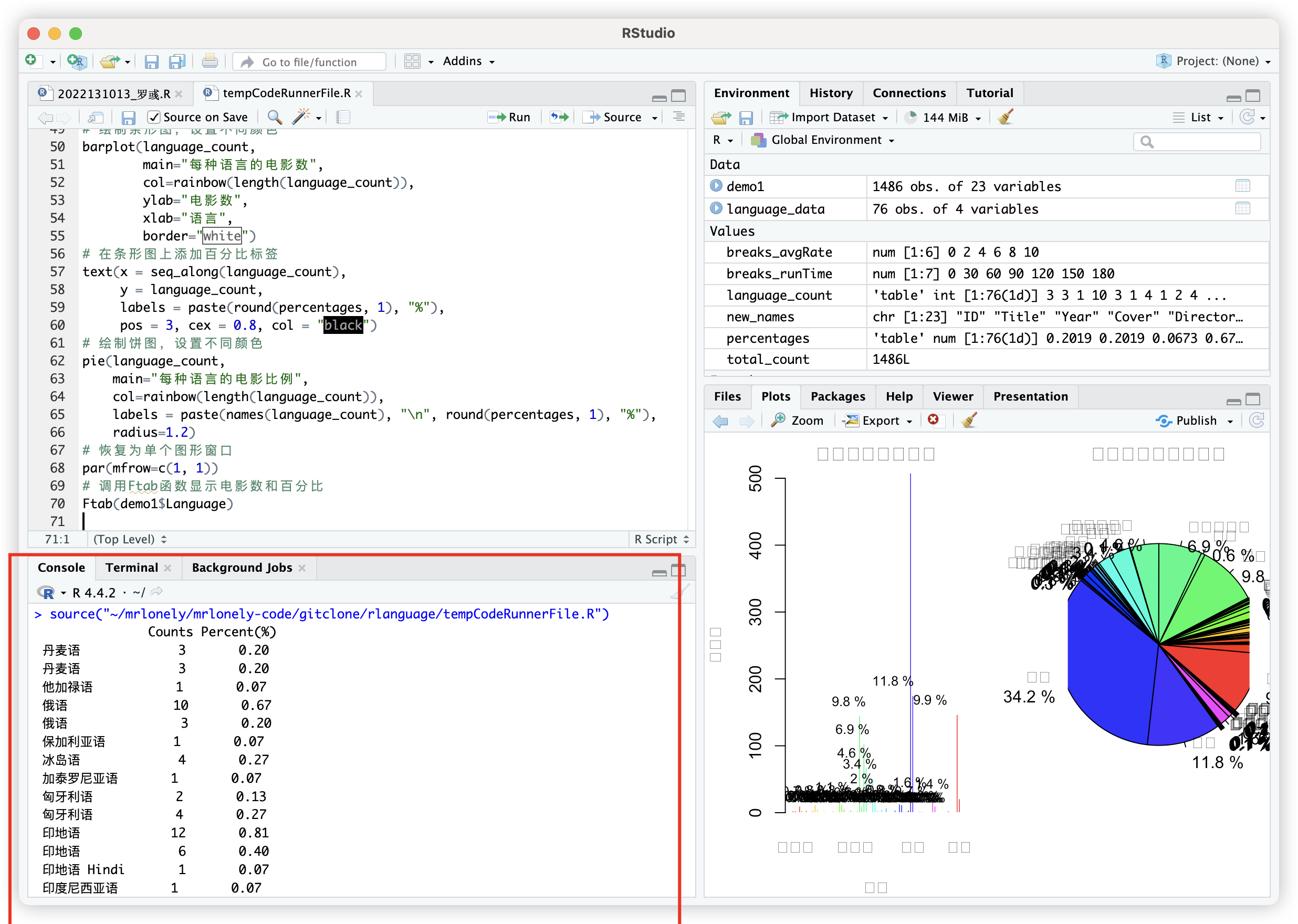
描述已自动生成2）. 结合下面的代码和图片，分析观察评分与评分人数、评分与五星人数之间的关系，并依据评分分组数据进行显示。



1. . 使用table函数统计每种语言的电影数，并使用barplot，pie绘制图形（左右排列），按不同颜色（非默认）区分。

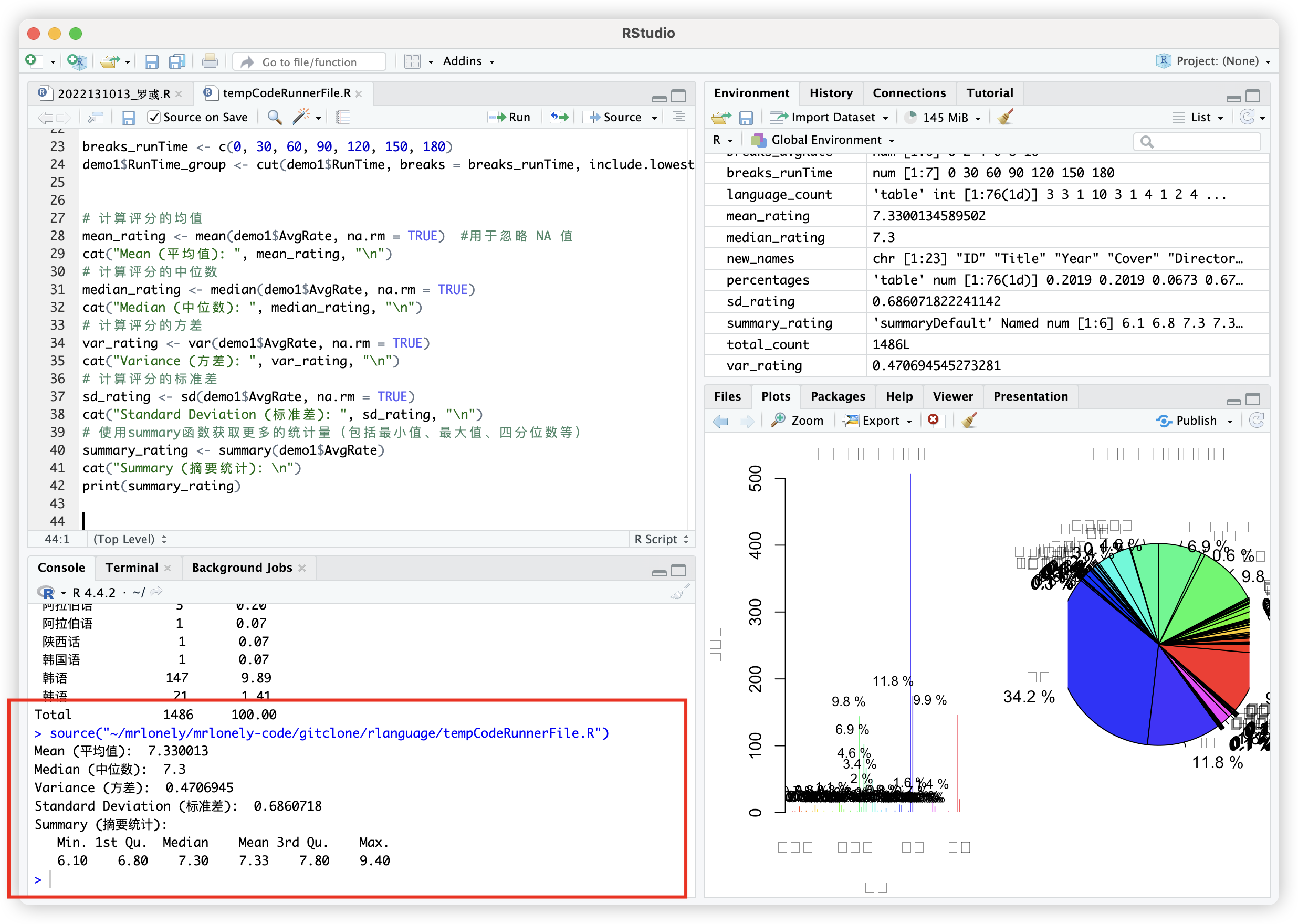


1. . 针对1生成的数据，求解每种语言的电影数占比，并将生成的数据和1中的数据左右拼接。



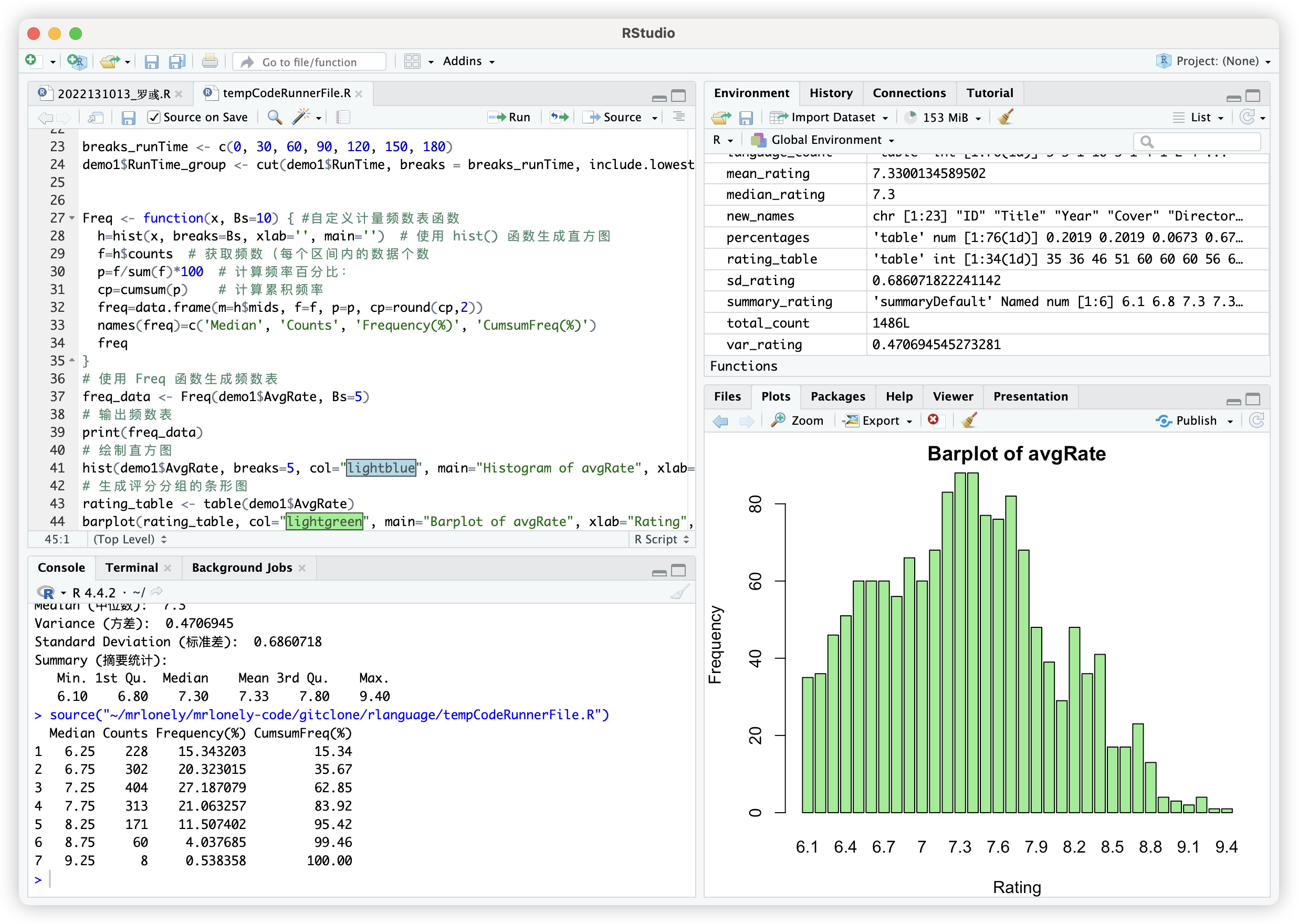
图形用户界面, 文本

描述已自动生成5）. 试分用mean, median, Var, sd, summary函数求评分变量的相关统计量。



图形用户界面, 应用程序, Word

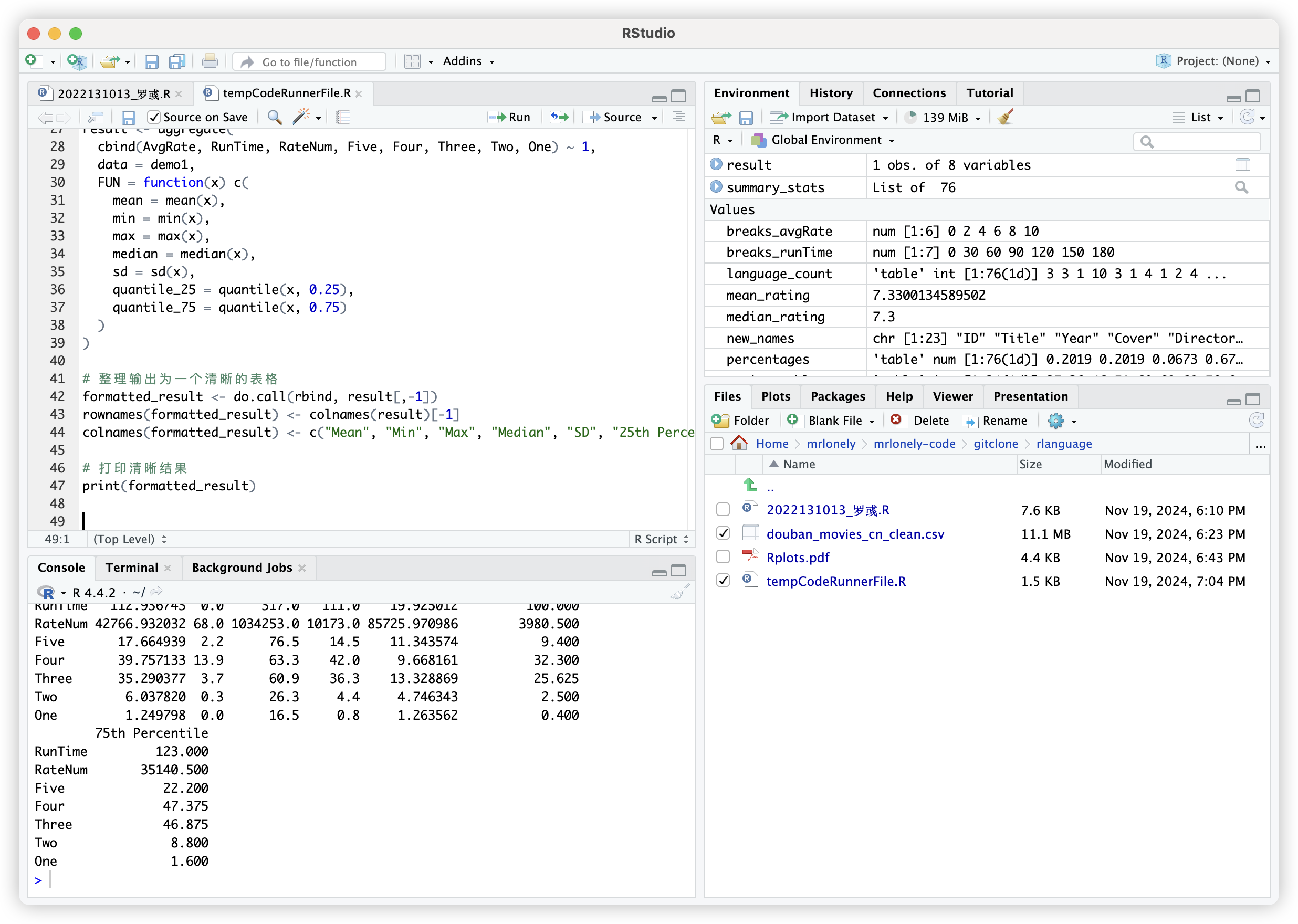
描述已自动生成6）. 试使用下面的函数生成相应数据。将函数生成的直方图与评分分组变量数据的barplot进行对比（相同的分组数目），观察它们的区别和联系。



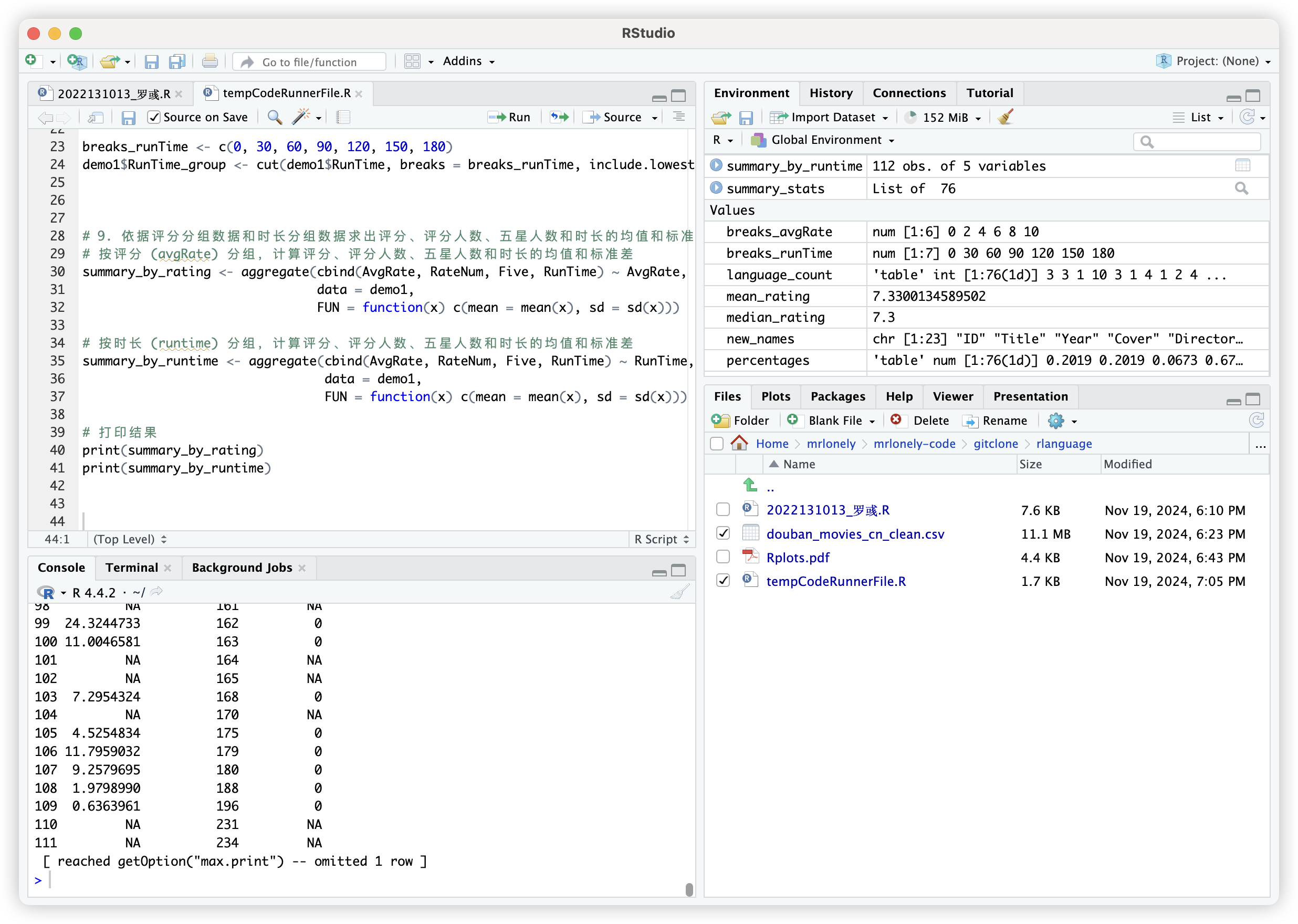
1. . 按照不同的语言分类得出评分数据的的基本统计量（summary函数）和boxplot箱式图和stripchart点带图（对角线排列）。



1. . 使用聚合函数求出连续型变量（包括：评分，时长，评分人数，五星、四星、三星、二星、一星）的基本统计量。



1. . 依据评分分组数据和时长分组数据求出评分、评分人数和时长的均值和标准差。



**四、实验结果：**

1. 完成了什么实验？

R语言的数据分析实验

2. 实验的结论。

R对数据分析很好用，但是完全可以被Python取代的

3. 遇到了什么困难，怎么解决的？

有一些环境问题， 最后直接使用rstudio来解决的；

4. 对实验有何认识？

这个实验很好可以增强我对R的认识